4. Linear Algebra

4.1. Vector Spaces and Subspaces

Definition 1. A *vector space* (or *linear space*) V over a field[1] F consists of a set on which two operations (called *addition* and *scalar multiplication*, respectively) are defined so that:

(a) (closed under addition)

$$\text{there exists a unique } \mathbf{u}+\mathbf{v} \in V \text{ for all } \mathbf{u},\mathbf{v} \in V;^{2}$$

(b) (closed under scalar multiplication)

$$\text{there exists a unique } c\mathbf{u} \in V \text{ for all } \mathbf{u} \in V.$$

and such that the following eight properties hold:

(i) (commutativity of addition).

$$\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}.$$

(ii) (associativity of addition).

$$(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w}).$$

(iii) (identity element of addition). There exists an element $\mathbf{0} \in V$, called the *zero vector*, such that

$$\mathbf{u}+\mathbf{0} = \mathbf{u} \text{ for all } \mathbf{u} \in V.$$

(iv) (inverse elements of addition). For every $\mathbf{u} \in V$, there exists an element $-\mathbf{u} \in V$, called the *additive inverse* of $\mathbf{u}$, such that

$$\mathbf{u}+(-\mathbf{u}) = \mathbf{0}.$$

(v) (compatibility of scalar multiplication with field multiplication).

$$(ab)\mathbf{u} = a(b\mathbf{u}).$$

[1] A field is a set together with two operations, called *addition* and *multiplication* respectively such that a few axioms hold. Two important examples are the real numbers R and the complex numbers C. We restrict our attention to the field of real numbers R in this text. For a formal definition of field, you are referred to Appendix C, *Friedberg, Insel and Spence (2002), Linear Algebra*.

[2] To emphasize that we are now living in vector spaces which differ from the metric spaces in previous chapters, we shall denote the elements (vectors) by $\mathbf{u},\mathbf{v}$ other than $\mathbf{x},\mathbf{y}$.

(vi) (identity element of scalar multiplication). There exists an element $1 \in F$, called the *multiplicative identity* in F, such that

$$1\mathbf{u} = \mathbf{u}.$$

(vii) (distributivity of scalar multiplication with respect to vector addition).

$$a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v}.$$

(viii) (distributivity of scalar multiplication with respect to field addition).

$$(a+b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}.$$

Example 1. The set of all k-tuples with entries from a field F is denoted by F^k. This set is a vector space over F with the operations of coordinatewise addition and scalar multiplications: that is, if $\mathbf{u} = (u_1, u_2, \cdots, u_k) \in F^k$, $\mathbf{v} = (v_1, v_2, \cdots, v_k) \in F^k$, and $c \in F$, then

$$\mathbf{u}+\mathbf{v} = (u_1+v_1, u_2+v_2, \cdots, u_k+v_k) \text{ and } c\mathbf{u} = (cu_1, cu_2, \cdots, cu_k).$$

It can be easily checked that the eight properties hold. Let F be the real field R, thus euclidean spaces R^k become vector spaces.

Example 2. The set of all $m \times n$ matrices with entries from a field F is a vector space, which we denote by $\mathbf{M}_{m\times n}(F)$, with the following operations of matrix addition and scalar multiplication: For $A, B \in \mathbf{M}_{m\times n}(F)$ and $c \in F$,

$$(A+B)_{ij} = A_{ij} + B_{ij} \text{ and } (cA)_{ij} = cA_{ij}$$

for $1 \le i \le m$ and $1 \le j \le n$.

Example 3. Let

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

and

$$g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$$

be polynomials with coefficients from a field F. Suppose that $m \le n$, and define $b_{m+1} = b_{m+2} = \cdots = b_n = 0$. Then $g(x)$ can be written as

$$g(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$$

Define

$$f(x) + g(x) = (a_n+b_n)x^n + (a_{n-1}+b_{n-1})x^{n-1} + \cdots + (a_1+b_1)x + (a_0+b_0)$$

and for any $c \in F$, define

$$cf(x) = ca_n x^n + ca_{n-1} x^{n-1} + \cdots + ca_1 x + ca_0.$$

With these operations of addition and scalar multiplication, the set of all polynomials with coefficients from F is a vector space, which we denote by $\mathbf{P}(F)$.

Exercise 4. Let $S = \{(a_1, a_2) : a_1, a_2 \in R\}$. For $(a_1, a_2), (b_1, b_2) \in S$ and $c \in R$, define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 - b_2) \text{ and } c(a_1, a_2) = (ca_1, ca_2).$$

Is S a vector space over R with these operations? Justify your answer.

Remark. Observe that in describing a vector space, it is necessary to specify not only the vectors but also the operations of addition and scalar multiplication.

Definition 2. A subset W of a vector space V over a field F is called a *subspace* of V if W is a vector space over F with the operations of addition and scalar multiplication defined on V.

Fortunately it is not necessary to verify all of the vector space properties to prove that a subset is a subspace. In fact, properties (i), (ii), (v)∼(viii) hold for a generic subset of vector space V and property (iv) is redundant given the following three conditions in Theorem 1.

Theorem 1. Let V be a vector space and W a subset of V. Then W is a subspace of V if and only if the following three properties hold for the operations defined in V.
(a) $\mathbf{0} \in W$.
(b) $\mathbf{u} + \mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$.
(c) $c\mathbf{u} \in W$ for all $c \in F$ and $\mathbf{u} \in W$.

Example 5. The set W of all symmetric matrices in $\mathbf{M}_{n \times n}(F)$ is a subspace of $\mathbf{M}_{n \times n}(F)$ since the conditions of Theorem 1 hold.

Theorem 2. Any intersection of subspaces of a vector space V is a subspace of V.

4.2. Linear Combinations and Systems of Linear Equations

Some economics models have a natural linear structure, while much more are described by a system of nonlinear equations. In the latter case, one applies Taylor theorem to convert them into an approximating linear system. By studying the properties of the linear system, we can learn a lot about the underlying nonlinear system. We'll take the above as the point of departure, and focus on solving systems of linear equations as follows.

Definition 3. Let V be a vector space over a field F. For some sets of vectors $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}$, a vector $\mathbf{v} \in V$ is called a *linear combination* of those vectors if there exist scalars $a_1, a_2, \cdots, a_n$ in F such that $\mathbf{v} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + \cdots + a_n\mathbf{u_n}$. In this case we call $a_1, a_2, \cdots, a_n$ the coefficients of the linear combination.

Example 6. Let $\mathbf{P}(R)$ be the vector space of all polynomials with coefficients taken from the real field R. Consider the vectors (polynomials) $p_0 = 1, p_1 = x, p_2 = x^2, \cdots, p_n = x^n$. Then any vector (polynomial) $p \in \mathbf{P}(R)$ is a linear combination of those vectors:

$$p = a_n p_n + a_{n-1} p_{n-1} + \cdots + a_1 p_1 + a_0 p_0. \qquad \triangle$$

In many situations, it is necessary to determine whether or not a vector can be expressed as a linear combination of other vectors, and if so, how. This question often reduces to the problem of solving a system of linear equations.

Exercise 7. For each of the following lists of vectors in R^3, determine whether the last vector can be expressed as a linear combination of the rest, and if so, how.

1. $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}.$
2. $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}.$ $\triangle$

Generally, we can rewrite any system of linear equations into matrix form and apply a technique called *Gaussian elimination* to solve it.

The system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is called a system of m linear equations in n unknowns.

The $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called the *coefficient matrix* of the system.

If we let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

then the system may be rewritten as a single matrix equation

$$A\mathbf{x} = \mathbf{b}$$

A solution to the system is an n-tuple

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \in F^n$$

such that $A\mathbf{s} = \mathbf{b}$. The set of all solutions to the system is called the *solution set*[3] of the system. The system is called *consistent* if its solution set is nonempty; otherwise it is called *inconsistent*.

We shall now consider the matrix $(A|b)$, the so-called *augmented matrix of the system* $A\mathbf{x} = \mathbf{b}$. Gaussian elimination involves applying elementary row operations to the augmented matrix.

[3]If its solution set is nonempty, then it might exists a unique solution or infinite solutions. Regarding the theoretical aspects of solution sets, you are referred to pages 168-175 (especially Theorem 3.8-3.9), *Friedberg, Insel and Spence (2002), Linear Algebra*.

Definition 4. Let A be an $m \times n$ matrix. Any one of the following three operations on the rows of A is called an *elementary row operation*:
(1) interchanging any two rows of A;
(2) multiplying any row of A by a nonzero scalar;
(3) adding any scalar multiple of a row of A to another row.

Theorem 3. Let $A\mathbf{x} = \mathbf{b}$ be a system of m linear equations in n unknowns. If $(A'|b')$ is obtained from $(A|b)$ by a finite number of elementary row operations, then the system $A'\mathbf{x} = \mathbf{b}'$ has the same solution set as the original system.

We now describe the procedure of Gaussian elimination[4] by solving Exercise 7(1) as follows:

$$(A|b) = \begin{pmatrix} 1 & 2 & -2 \\ 3 & 4 & 0 \\ 0 & -1 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & -2 & 6 \\ 0 & -1 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -3 \\ 0 & -1 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix} = (A'|b')$$

The result is the desired reduction of the augmented matrix $(A'|b')$, which corresponds to the new system of linear equations

$$x_1 + 0 = 4$$
$$0 + x_2 = -3$$

Thus the solution set is $\mathbf{s} = [4, -3]^T$. That is,

$$\begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}.$$

Exercise 7' Redo Exercise 7(2) using Gaussian elimination if you haven't.

Theorem 4. Let $A\mathbf{x} = \mathbf{b}$ be a system of linear equations. Then the system is consistent if and only if rank(A)=rank($A|b$).

Definition 5. Let $S = \{\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}\}$ be a nonempty subset of a vector space V. The *span* of S, denoted span(S), is the set consisting of all linear combinations of the vectors in S.

[4]For a detailed description of the procedure, readers are referred to pages 183-185, *Friedberg, Insel and Spence (2002), Linear Algebra*.

Example 8. In R^3, consider the set $S = \{(1,0,0),(0,1,0),(1,1,0)\}$. Then $span(S) = \{(a,b,0):$ for any $a,b \in R\}$. Thus the span of S contains all the points in the xy-plane. In this case, the span of the set is s subspace of R^3. This fact is true in general.

Theorem 5. The span of any subset S of a vector space V is a subspace of V. Moreover, any subspace of V that contains S must also contain the span of S.

Remark. We also say that the vectors in S *generate* the subspace $W = span(S)$.

4.3. Linear Independence, Basis and Dimension

Suppose V is a vector space and W is a subspace of V. It is always desirable to find a "smallest" finite subset S that generates W because we can then describe each vector in W as a linear combination of the smallest number of vectors in S. Consider, in Example 8, the subspace $W = span(S)$ can be generated by $S' = \{(1,0,0),(0,1,0)\}$. Generally, finding the smallest set S', however, involves the concept of linear independence.

Definition 6. A subset $S = \{\mathbf{u_1},\mathbf{u_2},\cdots,\mathbf{u_n}\}$ of a vector space V is called *linearly dependent* if there exist scalars $a_1,a_2,\cdots,a_n \in F$, not all zero, such that

$$a_1\mathbf{u_1} + a_2\mathbf{u_2} + \cdots + a_n\mathbf{u_n} = 0 \tag{1}$$

A subset $S = \{\mathbf{u_1},\mathbf{u_2},\cdots,\mathbf{u_n}\}$ of a vector space that is not linearly dependent is called *linearly independent*. In this case, equation (1) will imply $a_1 = a_2 = \cdots = a_n = 0$.

Exercise 9. Determine whether the vectors are linearly independent or not.

1. $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}.$
2. $\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$

Remark. If a set S of vectors is linearly dependent then there exists some vector $\mathbf{v} \in S$ can be expressed as a linear combination of the other vectors in S, and the subset obtained by removing $\mathbf{v}$ from S has the same span as S. It follows that if no proper subset of S generates the span of S, then S must be linearly independent.

Definition 7. A *basis* β for a vector space V is a linearly independent subset of V that generates V.

Theorem 6. Let V be a vector space and $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}\}$ be a subset of V. Then β is a basis for V if and only if each $\mathbf{v} \in V$ can be uniquely expressed as a linear combination of vectors of β, that is, can be expressed in the form

$$\mathbf{v} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + \cdots + a_n\mathbf{u_n}$$

for unique scalars $a_1, a_2, \cdots, a_n$.

Remark. Consider a generic vector space V and a basis $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}$. The theorem shows that each $\mathbf{v} \in V$ determines a unique n-tuple of scalars $a_1, a_2, \cdots, a_n$ and, conversely, each n-tuple of scalars determines a unique vector $\mathbf{v} \in V$ by using the entries of the n-tuple as the coefficients of a linear combination of $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}$. This fact suggests that V is like the vector space R^n, where n is the number of vectors in the basis for V.

The most familiar example of a basis is the set $\{e_1, e_2, \cdots, e_n\}$, where e_i is the vector in R^n whose i-th coordinate is 1 and other coordinates are 0. For each $\mathbf{v} = (v_1, v_2, \cdots, v_n) \in R^n$, it can be expressed as $\mathbf{v} = \sum v_i e_i$. We call $\{e_1, e_2, \cdots, e_n\}$ the *standard basis* of R^n.

Definition 8. Let V be a vector space having a finite basis. Then every basis for V contains the same number of vectors. The unique number of vectors in each basis for V is called the *dimension* of V and is denoted by dim(V).

Exercise 10. Consider the set of vectors

$$S = \{(2,-3,1), (1,4,-2), (-8,12,-4), (1,37,-17), (-3,-5,8)\}.$$

Let V be the vector space spanned by S. What is the dimension of V? Find a subset of S that is a basis for V.

4.4. Linear Transformations and Their Matrix Representations

4.4.1. Linear Transformations

So far, we developed the theory of abstract vector spaces. It is now natural to consider those functions defined on vector spaces. The most important class of such functions are called *linear transformations*, which in some sense "preserve" the structure of vector space.

Definition 9. Let V and W be vector spaces over F. We call a function $T: V \to W$ a *linear transformation from V to W* if, for all $\mathbf{u}, \mathbf{v} \in V$ and $c \in F$, we have

(a) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;
(b) $T(c\mathbf{u}) = cT(\mathbf{u})$.

We also define the *identity transformation* $I_V : V \to V$ by $I_V(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in V$ and the *zero transformation* $T_0 : V \to W$ by $T_0(\mathbf{u}) = \mathbf{0}$ for all $\mathbf{u} \in V$. They are two very important examples of linear transformations deserve their own notation.

Exercise 11. Let $V = C(R)$, the vector space of continuous real-valued functions on R. Let $a, b \in R, a < b$. Define $T: V \to R$ by

$$T(f) = \int_a^b f(t)dt$$

for all $f \in V$. Show that T is a linear transformation.

Example 12. For any angle θ, define $T_\theta : R^2 \to R^2$ by the rule: $T_\theta(u_1, u_2)$ is the vector obtained by *rotating* (u_1, u_2) counterclockwise by θ.
An explicit formula for T_θ can be determined. Let α be the angle that (u_1, u_2) makes with the positive x-axis, and let $r = \sqrt{u_1^2 + u_2^2}$. Then $u_1 = r\cos\alpha$ and $u_2 = r\sin\alpha$. It follows that

$$\begin{aligned}
T_\theta(u_1, u_2) =& T_\theta(r\cos\alpha, r\sin\alpha) \\
=& (r\cos(\alpha+\theta), r\sin(\alpha+\theta)) \\
=& (r\cos\alpha\cos\theta - r\sin\alpha\sin\theta, r\cos\alpha\sin\theta + r\sin\alpha\cos\theta) \\
=& (u_1\cos\theta - u_2\sin\theta, u_1\sin\theta + u_2\cos\theta).
\end{aligned}$$

It is now easy to show that T_θ is linear.

Remark. The applications of linear algebra to geometry are wide and varied. The main reason for this is that most of the important geometrical transformations are linear, for instance, rotation, reflection ($T(u_1, u_2) = (u_1, -u_2)$), and projection ($T(u_1, u_2) = (u_1, 0)$).

4.4.2. The Matrix Representation of a Linear Transformation

We now turn to the most useful approach to the analysis of a linear transformation on a finite-dimensional vector space: the representation of a linear transformation by a matrix. This is done by applying the concept of an *ordered basis* for a vector space.

Definition 10. Let V be a finite-dimensional vector space. An *ordered basis* for V is a basis for V endowed with a specific order.

Recall the remark under Theorem 5, we can identify abstract vectors in an n-dimensional vector space with n-tuples.

Definition 11. Let $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_n}\}$ be an ordered basis for a finite-dimensional vector space V. For $\mathbf{u} \in V$, let $a_1, a_2, \cdots, a_n$ be the unique scalars such that

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{u_i}$$

We define the *coordinate vector of* $\mathbf{u}$ *relative to* β, denoted $[\mathbf{u}]_\beta$, by

$$[\mathbf{u}]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Example 13. Let $V = \mathbf{P_2}(R)$, and let $\beta = \{1, x, x^2\}$ be the standard ordered basis for V. If $f(x) = 4 + 6x - 7x^2$, then

$$[f]_\beta = \begin{pmatrix} 4 \\ 6 \\ -7 \end{pmatrix}$$

Definition 12. Suppose that V and W are finite-dimensional vector spaces with ordered bases $\beta = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$ and $\gamma = \{\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_m}\}$, respectively. Let $T : V \to W$ be linear. Then for each j, $1 \le j \le n$, there exist unique scalars $a_{ij} \in F$, $1 \le i \le m$, such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{for } 1 \le j \le n.$$

We call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$ the *matrix representation of T in the ordered bases β and γ* and write $A = [T]_\beta^\gamma$. If $V = W$ and $\beta = \gamma$, then we write $A = [T]_\beta$.

Example 14. Revisit Example 12. Let β be the standard ordered basis for R^2. Now

$$T_\theta(1,0) = (\cos\theta, \sin\theta)$$

and

$$T_\theta(0,1) = (-\sin\theta, \cos\theta).$$

Hence

$$[T]_\beta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Exercise 15. Let $T: R^2 \to R^3$ be the linear transformation defined by

$$T(u_1, u_2) = (u_1 + 3u_2, 0, 2u_1 - 4u_2).$$

Let β and γ be the standard ordered bases for R^2 and R^3, respectively. Find the matrix representation for this linear transformation.

The next result justifies much of our past work. It utilizes the matrix representation of a linear transformation to evaluate the transformation at any given vector.

Theorem 7. Let V and W be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let $T: V \to W$ be linear. Then, for each $\mathbf{u} \in V$, we have

$$[T(\mathbf{u})]_\gamma = [T]_\beta^\gamma [\mathbf{u}]_\beta.$$

Example 16. Revisit Example 14. Let $T_\theta: R^2 \to R^2$ be the counterclockwise rotation by angle $\theta = \frac{\pi}{2}$, and let β be the standard ordered basis. From Example 15, we have

$$[T]_\beta = \begin{pmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

We illustrate Theorem 7 by verifying that $[T(\mathbf{u})]_\beta = [T]_\beta [\mathbf{u}]_\beta$, where $\mathbf{u} \in R^2$ is the vector $\mathbf{u} = (1,1)^T$. From geometry, we have

$$[T(\mathbf{u})]_\beta = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

but also

$$[T]_\beta [\mathbf{u}]_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Now that we have defined a procedure for associating matrices with linear transformations, we show below that this association "preserves" the structure of linear transformations.

Theorem 8. Let V and W be vector spaces over a field F, and let $T, U: V \to W$ be linear. We define:
(a). (addition). $T + U: V \to W$ by $(T+U)(\mathbf{v}) = T(\mathbf{v}) + U(\mathbf{v})$ for all $\mathbf{v} \in V$,
(b). (scalar multiplication). $aT: V \to W$ by $(aT)(\mathbf{v}) = aT(\mathbf{v})$ for all $\mathbf{v} \in V$. With the operations of addition and scalar multiplication, the collection of all linear transformations from V to W forms a vector space over F, denoted by $\mathscr{L}(V, W)$. In the case that $V = W$, we write $\mathscr{L}(V)$.

Theorem 9. Let V and W be finite-dimensional vector spaces with ordered bases $\beta = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$ and $\gamma = \{\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_m}\}$, respectively, and let $T, U : V \to W$ be linear transformations. Then their matrix representations satisfy:
(a) $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$,
(b) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ for all scalars a.
And, the collection of all matrices representing the linear transformations forms a vector space $\mathbf{M}_{m\times n}$.

Remark. Compare Theorem 7 with Example 2. In the spirit of Theorem 7, we shall regard matrices as linear transformations, and $\mathbf{M}_{m\times n}$ as $\mathscr{L}(V,W)$.

4.4.3. Composition of Linear Transformations and Matrix Multiplication

The question now arises as to how the matrix representation of a composite of linear transformations is related to the matrix representation of each of the associated linear transformations. It is the attempt to answer this question that leads to a definition of *matrix multiplication.*

Theorem 10. Let V, W and Z be vector spaces over the same field F, and let $T : V \to W$ and $U : W \to Z$ be linear. We define the *composite of linear transformations* $UT : V \to Z$ by $(UT)(\mathbf{v}) = U(T(\mathbf{v}))$ for all $\mathbf{v} \in V$. Then UT is linear.

To define the product of two matrices, let $T : V \to W$ and $U : W \to Z$ be linear. Let $A = [U]_\beta^\gamma$ and $B = [T]_\alpha^\beta$ where $\alpha = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$, $\beta = \{\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_m}\}$ and $\gamma = \{\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_p}\}$ are ordered bases for V, W and Z, respectively. For $1 \le j \le n$ we have

$$\begin{aligned}
(UT)(v_j) = U(T(v_j)) = U(\sum_{k=1}^m B_{kj} w_k) &= \sum k = 1^m B_{kj} U(w_k) \\
&= \sum_{k=1}^m B_{kj} \left(\sum_{i=1}^p A_{ik} z_i \right) = \sum_{i=1}^p \left(\sum_{k=1}^m A_{ik} B_{kj} \right) z_i \\
&= \sum_{i=1}^p C_{ij} z_i,
\end{aligned}$$

where

$$C_{ij} = \sum_{k=1}^m A_{ik} B_{kj}.$$

This computation motivates the following definition of generic matrix multiplication.

Definition 13. Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We define the *product* of A and B, denoted AB, to be the $m \times p$ matrix such that

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \quad \text{for } 1 \leq i \leq m, 1 \leq j \leq p.$$

Remark. Note that $(AB)_{ij}$ is the sum of products of corresponding entries from the ith row of A and jth column of B. If we decompose matrix A into its row vectors $\mathbf{a_i}$, and the matrix B into its column vectors $\mathbf{b_i}$:

$$A = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_n} \end{pmatrix}, \quad B = (\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_p})$$

Then:

$$AB = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_n} \end{pmatrix} (\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_p}) = \begin{pmatrix} (\mathbf{a_1} \cdot \mathbf{b_1}) & (\mathbf{a_1} \cdot \mathbf{b_2}) & \cdots & (\mathbf{a_1} \cdot \mathbf{b_p}) \\ (\mathbf{a_2} \cdot \mathbf{b_1}) & (\mathbf{a_2} \cdot \mathbf{b_2}) & \cdots & (\mathbf{a_2} \cdot \mathbf{b_p}) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{a_n} \cdot \mathbf{b_1}) & (\mathbf{a_n} \cdot \mathbf{b_2}) & \cdots & (\mathbf{a_n} \cdot \mathbf{b_p}) \end{pmatrix}$$

The next theorem is an immediate consequence of the definition of matrix multiplication.

Theorem 11. Let V, W and Z be finite-dimensional vector spaces with ordered bases α, β and γ, respectively. Let $T : V \to W$ and $U : W \to Z$ be linear. Then

$$[UT]_{\alpha}^{\gamma} = [U]_{\beta}^{\gamma}[T]_{\alpha}^{\beta}$$

Example 17. Revisit Example 14. Let $U_\theta : R^2 \to R^2$ and $T_\theta : R^2 \to R^2$ both be the counter-clockwise rotation by angel θ. From geometry, it follows that $UT_{2\theta} = U_\theta T_\theta$ is the counter-clockwise rotation by angle 2θ. Thus, Example 15 implies that

$$[UT]_\beta = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}$$

To illustrate Theorem 9, observe that

$$\begin{aligned} [UT]_\beta = [U]_\beta [T]_\beta &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{pmatrix} = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} \end{aligned}$$

We complete this section with some basic properties of matrix operations.

Theorem 12. Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$ matrices. Then
(a) $A(B+C) = AB+AC$ and $(D+E)A = DA+EA$.
(b) $a(AB) = (aA)B = A(aB)$ for any scalar a.
(c) $I_m A = A = AI_n$.
(d) $(AB)^T = B^T A^T$.

Remark. As in the case with composition of functions, we have that matrix multiplication is not commutative, that is, it need not be true that $AB = BA$.

4.5. The Rank of a Matrix and Matrix Inverses

Definition 14. Let V and W be vector spaces. If $T \in \mathscr{L}(V,W)$ and $A = [T]_\beta^\gamma$ where β and γ are ordered bases of V and W. We define the *range* $R(T)$ of T to be the subset of W consisting of all images (under T) of vectors in V; that is, $R(T) = \{T(\mathbf{u}) : \mathbf{u} \in V\}$. And, we define the *rank* of T and also the rank of A, denoted rank(T) and rank(A), respectively, to be the dimensions of $R(T)$.

Theorem 13. Let V and W be vector spaces, and let $T : V \to W$ be linear. If $\beta = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is a basis for V, then

$$R(T) = span(T(\beta)) = span(\{T(\mathbf{e}_1), T(\mathbf{e}_2), \cdots, T(\mathbf{e}_n)\}).$$

Remark. Note that $[T(\mathbf{e}_i)]_\beta = A[\mathbf{e}_i]_\beta$, which gives the ith column of matrix A. Therefore, it is clear that the rank of matrix A is equal to the dimension of the subspace spanned by column vectors of A, that is, the maximum number of *linearly independent column vectors* of A.

Theorem 14. Let A be an $m \times n$ matrix. Then
(a) rank(A^T) = rank(A).
(b) The rank of any matrix equals the maximum number of its linearly independent rows; that is, the rank of a matrix is the dimension of the subspace generated by its rows.
(c) The rows and columns of any matrix generate subspaces of the same dimension, numerically equal to the rank of the matrix.

The following theorem allows us to find the rank of a matrix.

Theorem 15. Elementary row and column operations on a matrix are rank-preserving.

Exercise 18. Find the rank for each of the following matrices.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \end{pmatrix}.$$

Theorem 16. Let A and B be matrices such that the product AB is defined. Then
(a) rank(AB)$\leq$ rank(A).
(b) rank(AB)$\leq$ rank(B).

Definition 15. Let V and W be vector spaces, and let $T : V \to W$ be linear. A function $U : W \to V$ is said to be an *inverse* of T if $TU = I_W$ and $UT = I_V$. If T has an inverse, then T is said to be *invertible*, and the inverse is unique and denoted by T^{-1}. The definition is analogous for their matrix representations A and B.

Remark. (a) A priori it might be possible for such a function U to exist when $m \neq n$, in which case a non-square matrix would have an inverse. However, this is not actually possible.[5] Therefore, the definition for invertibility is sometimes restricted to square matrix:

A square $n \times n$ matrix A is called *invertible* (or *nonsingular, nondegenerate*) if there exists an $n \times n$ matrix B such that $AB = BA = I_n$. In this case, the unique matrix B is called the *inverse* of A, denoted by A^{-1}.

(b) For a non-square $m \times n$ matrix A, it does not have an inverse. However, it may have a left inverse or right inverse:

If A has rank n, then A has a *left inverse*: an $m \times n$ matrix B such that $BA = I_n$. If A has rank m, then it has a *right inverse*: an $n \times m$ matrix B such that $AB = I_m$.

Theorem 17. Let A be a square $n \times n$ matrix. A in invertible if and only if A has full rank; that is, rank(A)=n.

The following theorem allows us to find the inverse of a matrix.

Theorem 18. If A is an invertible $n \times n$ matrix, then it is possible to transform the matrix $(A|I_n)$ into the matrix $(I_n|A^{-1})$ by means of a finite number of *elementary row operations*.

[5] For a formal answer to this question, readers are referred to Section 2.4 (especially Theorem 2.19 and the definition above it), *Friedberg, Insel and Spence (2002), Linear Algebra*.

Exercise 19. Determine whether the matrix

$$A = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix}$$

is invertible, and if it is, compute its inverse.

Theorem 19. Let A be an $m \times n$ matrix. If P and Q are invertible $m \times m$ and $n \times n$ matrices, respectively, then
(a) rank(AQ)=rank(A),
(b) rank(PA)=rank(A),
(c) rank(PAQ) = rank(A).

4.6. Determinants and Cramer's Rule

The role of determinant in the study of linear algebra is less central than in former times. As for our applications, it mainly appears in the Cramer's rule, computing eigenvalues and defining Jacobians in multivariate calculus.

There are various equivalent ways to define the determinant of a square matrix A. Perhaps the simplest way to express the determinant is by recursively considering *cofactors*.

Definition 16. Let A be an $n \times n$ matrix. Let M_{ij} be the matrix obtained by deleting the i-th row and j-th column of A. The *cofactor* of entry A_{ij}, denoted C_{ij}, is defined by

$$C_{ij} = (-1)^{i+j} det(M_{ij}).$$

Example 20. For

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

the cofactor of A_{23} is

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = 6.$$

Definition 17. Let A be an $n \times n$ matrix. If $n = 1$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as

$$det(A) = \sum_{j=1}^{n} A_{1j} C_{1j} = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} det(M_{1j}).$$

Exercise 21. For

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \\ 2 & 3 & 4 \end{pmatrix},$$

calculate det(A).

The following rules can be used to simplify the evaluation of a determinant.

Theorem 20. Let A be an $n \times n$ matrix.
(a) If B is a matrix obtained by interchanging any two rows of A, then det(B)=-det(A).
(b) If B is a matrix obtained by multiplying a row of A by a nonzero scalar k, then det(B)=k det(A).
(c) If B is a matrix obtained by adding a multiple of one row of A to another row of A, then det(B)=det(A).

Theorem 21. For any square matrix A, det(A^T)=det(A).

Remark. This means that you can also use *elementary column operations* to evaluate determinants as well.

Exercise 22. For

$$A = \begin{pmatrix} 2 & 1 & -1 & 2 \\ 3 & 0 & 0 & 1 \\ 2 & 1 & 2 & 0 \\ 3 & 1 & 1 & 2 \end{pmatrix},$$

calculate det(A).

Theorem 22. Some basic properties of determinants are:
(a) det(I_n)=1 where I_n is the $n \times n$ identity matrix.
(b) det(A^{-1}) = $\frac{1}{\det(A)}$ = det$(A)^{-1}$.
(c) For square matrices A and B of equal size, det(AB)=det(A)det(B).
(d) If A is a triangular matrix, i.e. $A_{ij} = 0$ whenever $i > j$ or, alternatively, whenever $i < j$, then its determinant equals the product of the diagonal entries:

$$det(A) = \prod_{i=1}^{n} A_{ii}.$$

Theorem 23. A square matrix A is invertible if and only if det$(A) \neq 0$.

Theorem 24. If A is an $n \times n$ with rank(A)<n, then det(A)=0.

Theorem 25. (Cramer's rule). Consider a system of n linear equations for n unknowns, represented in matrix multiplication form as follows:

$$A\mathbf{x} = \mathbf{b}$$

where the $n \times n$ matrix A has a nonzero determinant. Then the system has a unique solution, whose individual values for the unknowns are given by:

$$x_i = \frac{det(A_i)}{det(A)} \quad i = 1, 2, \cdots, n$$

where A_i is the matrix formed by replacing the i-th column of A by the column vector b.

Exercise 23. Solve the following system of equations using Cramer's rule.

$$\begin{aligned} 2x + y + z &= 3 \\ x - y - z &= 0 \\ x + 2y + z &= 0 \end{aligned}$$

4.7. Quadratic Forms and Definite Matrices

Definition 18. Let A be an $n \times n$ symmetric matrix with real entries and let $\mathbf{x}$ denote an $n \times 1$ column vector. Then $Q = \mathbf{x}'A\mathbf{x}$ is said to be a *quadratic form*. Note that

$$Q = \mathbf{x}'A\mathbf{x} = \sum_{i \leq j} a_{ij} x_i x_j$$

Example 24. Consider the matrix

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

and the vector $\mathbf{x}$. Q is given by

$$\begin{aligned} Q &= \mathbf{x}'A\mathbf{x} = (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= (x_1 + 2x_2, 2x_1 + x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + 2x_1x_2 + 2x_1x_2 + x_2^2 \\ &= x_1^2 + 4x_1x_2 + x_2^2 \end{aligned}$$

Remark. The matrix associated with a quadratic form B need not be symmetric. However, no loss of generality is obtained by assuming B is symmetric. Consider a nonsymmetric matrix B and define $A = \frac{1}{2}(B + B^T)$, A is now symmetric and $\mathbf{x}'A\mathbf{x} = \mathbf{x}'B\mathbf{x}$.

Definition 19. (Classification of the quadratic form) A quadratic form $Q = \mathbf{x}'A\mathbf{x}$ is said to be:
(a) positive definite: $Q > 0$ when $x \neq 0$.
(b) positive semidefinite: $Q \geq 0$ for all $\mathbf{x}$ and $Q = 0$ for some $x \neq 0$.
(c) negative definite: $Q < 0$ when $x \neq 0$.
(d) negative semidefinite: $Q \leq 0$ for all $\mathbf{x}$ and $Q = 0$ for some $x \neq 0$.
(e) indefinite: $Q \geq 0$ for some $\mathbf{x}$ and $Q \leq 0$ for some other $\mathbf{x}$.

Positive definiteness can be tested by using leading principal minors.

Definition 20. A k-th order *leading principal minor* of an $n \times n$ matrix A is defined as the determinant of matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix} \quad k \in \{1, 2, \cdots, n\}.$$

denoted by Δ_k. A *principal minor*, denoted B_k, is the determinant of a principal submatrix of A where a *principal submatrix* is a matrix formed from the matrix A by taking a subset consisting of k rows and column elements from the same numbered columns.

Theorem 26. Let A be an $n \times n$ real symmetric matrix. Then
(a) A is positive definite if and only if $\Delta_r > 0$ for $r = 1, 2, \cdots, n$.
(b) A is negative definite if and only if $(-1)^r \Delta_r > 0$ for $r = 1, 2, \cdots, n$.
(c) If some k-th order principal minor of A is non-zero but does not fit either the above two sign patterns, then A is indefinite.

Theorem 27. Let A be an $n \times n$ real symmetric matrix. Then
(a) A is positive semi-definite if and only if $B_r \geq 0$ for $r = 1, 2, \cdots, n$.
(b) A is negative semi-definite if and only if $(-1)^r B_r \geq 0$ for $r = 1, 2, \cdots, n$.

Exercise 25. Check the definiteness of the following matrices.

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}, B = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 4 \end{pmatrix}, D = \begin{pmatrix} 1 & 3 & 2 & 3 \\ 3 & 5 & -2 & 4 \\ 3 & 4 & 1 & 2 \\ 7 & 3 & 1 & 2 \end{pmatrix}.$$

Theorem 28. Let A be an $n \times n$ real symmetric matrix. Then
(a) A is positive definite if and only if all its eigenvalues are strictly positive.
(b) A is negative definite if and only if all its eigenvalues are strictly negative.
(a) A is positive semi-definite if and only if all its eigenvalues are weakly positive.
(a) A is negative semi-definite if and only if all its eigenvalues are weakly negative.

4.8. More Exercises

Exercise 26. Let $S = \{(a_1, a_2) : a_1, a_2 \in R\}$. For $(a_1, a_2), (b_1, b_2) \in S$ and $c \in R$, define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 b_2) \text{ and } c(a_1, a_2) = (ca_1, a_2).$$

Is S a vector space over R with these operations? Justify your answer.

Exercise 27. Consider the following list of vectors in R^3. Let V be the vector space spanned by them. What is the dimension of V? Find a basis for V.

$$\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$